1 Introduction

In this report we discuss the issues in document understanding in the context of postal address recognition. The recognition methodology we propose employs the techniques of mathematical morphology, spatial relational matching, and contextual decision theory to perform word matching. Our recognition methodology does not explicitly involve recognizing segmented characters. Rather, the morphologically determined features and primitives are used directly in word matching. A side effect of this word matching is character segmentation and recognition.

2 Word Recognition

Computer recognition of words and characters, machine-generated or hand-printed, is a complex procedure requiring a variety of steps that successively transform the iconic data to recognition information. We hypothesize that the failure of today's recognition technology to handle unconstrained hand-printed characters of script is due to the fact that the existing algorithms do not develop one or more of the necessary steps to a high enough degree. Our thesis is that there is no shortcut to achieving a high read rate and a low substitution error. A recognition methodology must pay substantial attention to each of the following five steps: conditioning, labeling, grouping, extracting, and matching. It must also be able to incorporate the performance characterization of the method employed in each step.

Consider a phenomenological analysis of what it means to recognize a character. Consider for a moment the nature of data input to a recognition algorithm and the nature of the data output from a recognition algorithm. Input to the algorithm is the digital data structure, a large array of non-negative numbers, each one of which, considered individually, is mostly irrelevant to the recognition process. Output from the algorithm is a symbol string representing the linear sequences of the characters of the identified word. Here, each character corresponds to some sub-array of the large input array. Thus the input image is composed of a sequence of sub-images which may even have some overlap. Existing character recognition algorithms consist of a segmentation step and an identification step. While these steps are necessary, there is no need for segmentation to be done prior to and independent of identification. This kind of problem decomposition can only work for the simplest of word recognition problems. For more general word
recognition, a context-based analysis is more appropriate.

The solution to the word recognition problem must indirectly involve some process that can indicate for each identified character the subset of the input data array that constitutes the character. Now consider how it is that a process can identify a character from iconic data. An iconic letter or number is recognizable because of its shape. In the iconic form there is information relative to a distinction between the background and the foreground, and the shape of the character is associated with the foreground. This being the case, we would expect a recognition methodology to provide a way of processing the data to make a distinction between background and foreground. Indeed, such a thresholding step is involved in almost every existing character recognition algorithm. Here, however, we must step and consider what makes a point in the input subarray part of the background or part of the foreground. The direct answer is that foreground points are black and background points are white. This, though, is not the correct answer. A foreground point is a foreground point because it is part of the character. Being part of the character depends on the local shape of the character near the point, the blackness of the point, and the relative blackness of other nearby points which are part of the shape. Just as there is no need for an independent segmentation phase to precede the identification phase, there is no need for a fixed uniform thresholding phase to precede the primitive identification phase.

Thus, the segmentation and thresholding steps which are inherent to almost every existing character recognition algorithm are, in general, inadequate. We believe that this is why the existing algorithms have not been able to perform well.

The methodology we propose to use to overcome these inadequacies is one in which the parameters for the thresholding and segmentation steps are linked back to the word recognition step in two ways. In the first way the segmentation is a direct part of the recognition instead of preceding the recognition. In the second way, the algorithm will have an iterative closed loop form with the recognition phase driving the final choice of recognition parameter settings. In order for a control structure to be able to redefine parameter settings for earlier algorithm steps on the basis of outputs from later algorithm steps, each algorithm step must have associated with it a performance characterization. Performance characterization of an algorithm step is like a complex version of a reference operating curve for a mechanical or electrical device.
3 Methodology and Performance Characterization

The recognition methodology we propose employs the techniques of mathematical morphology, spatial relational matching, and contextual decision theory to perform word matching. Our recognition methodology does not explicitly involve recognizing segmented characters. Rather, the morphologically determined features and primitives are used directly in word matching. A side effect of this word matching is character segmentation and recognition. The algorithm will have an iterative closed loop form with the recognition phase driving the final choice of the parameter setting for the intermediate stages. First, noise cleaning and normalization will be performed in a preprocessing step. Next, distinguishing primitive features in the image and their spatial relations will be detected by morphological processing. These primitives and their relationships will then be matched against the features and relations of the candidate words in the lexicon. In the context of the already-recognized fields, the number of words that could possibly match the image word will be greatly reduced. A set of planned, controlled experiments will be performed to characterize the performance of the overall system as well as to find its optimal operating point.

The morphological processing used in our system to detect various features in the image and their spatial relationships will be similar to the one used by Haralick in [KH90]. The spatial, relational matching problem will be posed in the framework developed by Haralick and Shapiro[HS79, HS80, Har89a, SH81]. A formal theory for contextual decision making, similar to Haralick’s work in [Har83, HJ86], will be used to make use of the dynamically changing context in the system. Finally, the performance of each stage in the algorithm, as well as of the complete system, will be characterized in terms of the various parameters of the algorithm. Design of experimental protocol is treated by Haralick in [Har89c, Har89b, KJHP90, VH92].

4 Summary

We claim that the failure of current document understanding and vision systems is due to an open loop approach to the research problem. In contrast, we proposed a closed loop approach to the document understanding problem and takes performance evaluation of each subsystem and context into consideration.
References


